Europäisches Patentamt | Eur pean Patent Office | Office européen des brevets

EP 99/00860

Bescheinigung | Certificate | Attestation

5

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

98300952.3

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts:
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

M.B. RIJLING

DEN HAAG, DEN
THE HAGUE,
LA HAYE, LE 06/04/99

Europäisches Patentamt
Eur pean Patent Office
Office eur péen des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.:
Demande n°: 98300952.3

Anmeldetag:
Date of filing:
Date de dépôt: 10/02/98

Anmelder:
Applicant(s):
Demandeur(s):
GIST-BROCADES B.V.
2600 MA Delft
NETHERLANDS

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Endo-xylogalacturonase and homologues thereof

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat: State: Pays:	Tag: Date: Date:	Aktenzeichen: File no. Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/[redacted]/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

(2)

G-B Ref: EUR-2895P/Kemp Ref: N.73992

ENDO-XYLOGALACTURONASE AND HOMOLOGUES THEREOF

Field of the Invention

The present invention relates to a novel endo-xylogalacturonase and homologues thereof. It further relates to the use of said novel endo-xylogalacturonase in a method of processing plant or pectin-containing material to produce fruit juice and other plant extracts.

Background to the Invention

Enzyme preparations are often used during the processing of plant materials, for example in the steps of extraction and liquefaction of fruit and fruit juice and their filtration and clarification. Commercial enzyme preparations contain a mixture of enzymes which degrade the pectin polymers which are a major component of plant cell walls. Such enzymes include pectin lyases, polygalacturonases, pectin esterases, cellulases, xyloglucanases, galactanases and arabinases.

Pectins occur in nature as constituents of higher plant cell walls. They are found in primary cell wall lamella where they are embedded in between the cellulose fibrils. The composition of pectin is variable among plant species and moreover dependent on the age and the maturity of the fruit. Among the richer sources of pectins are lemons and oranges, which can contain up to 30% of this polysaccharide. Most pectin polymers are comprised of 'smooth' homogalacturonan regions and ramified 'hairy' regions. The 'smooth' regions consist of a linear homogalacturonan backbone.

By contrast, the 'hairy' regions, as identified in apples, consist of three different subunits: subunit I is xylogalacturonan (a galacturonan backbone heavily substituted with xylose); subunit II is a short section of a rhamnogalacturonan backbone, rich in relatively long arabinan sidechains (the 'hairs'); subunit III is a rhamnogalacturonan oligomer, having a backbone consisting of an alternating sequence of rhamnose and galacturonic acid.

Many of the well-known pectinases degrade only the 'smooth' part of the pectin polymer leaving the 'hairy' regions intact. Consequently, during for example apple juice production, the non-degraded parts of the pectin polymer cause production losses due to inefficient filtration as a result of fouling of the ultrafiltration membrane.

Several enzymes have been reported which can degrade parts of the 'hairy' region, for example the rhamnogalacturonan regions of the backbone (subunit III). These enzymes are referred to as rhamnogalacturonases (RGases), of which there are several types. However, so far the xylogalacturonan part of the 'hairy' regions (subunit I) has been resistant to enzymatic digestion.

Since xylogalacturonan has also been found in many other plants, e.g. leguminous plants like soybeans and peas, watermelons, grapes, pine pollen enzymes to degrade this polymer are required for the processing of plant material.

Beldman *et al,* 1996, reports the identification of an exo-galacturonase that is not hindered by the single unit xylose side-chains and is able to degrade xylogalacturonan using a soluble 'hairy' pectic polysaccharide from soy as substrate. This enzyme acts in an exo-fashion yielding galacturonic acid or a disaccharide consisting of galacturonic acid and xylose. The enzyme was purified to near homogeneity and partially characterised. By contrast to known RGases, that do not degrade homogalacturonic acid, this enzyme is not very specific for xylogalacturonan as it also acts on pectic acid. In addition, this enzyme is not able to digest the xylogalacturonan backbone in a random fashion, and therefore to date there are no known enzymes possessing endo-xylogalacuronase activity.

Disclosure of the Invention

The present invention has resulted from the isolation and characterisation of a cDNA encoding a novel endo-xylogalacturonase. The endo-xylogalacturonase cDNA sequence is set out as SEQ. ID No. 1. The amino acid sequence of the ORF from nucleotides 98 to 1315 is set out as SEQ. ID No. 2.

In a first aspect of the invention there is provided a polypeptide possessing endo-xylogalacturonase activity. There is also provided a polypeptide comprising an endo-xylogalacturonase, such as a polypeptide comprising the sequence set out in SEQ ID No. 2, or a polypeptide substantially homologous thereto, or a fragment of the polypeptide of SEQ ID No. 2, said fragment having at least 5 amino acids.

The polypeptide of the invention preferably has one or more of the following additional features:

(1) possesses endo-xylogalacturonase activity;

(2) has an optimal pH-range from 3 to 6;

(3) has optimum activity at a temperature of from 50 to 70°C.

(4) a molecular weight (deglycosylated) of from 40 to 50 kDa.

"Endo-xylogalacturonase activity" is defined as the ability to cleave a galacturonic acid polymer, for example as found in pectin, which is at least partially substituted with xylose residues. Such cleavage preferably occurs at internal glycosidic bonds. More preferably said cleavage occurs at a [galacturonic acid (1-4) galacturonic acid] linkage. Preferably, said polypeptide does not cleave terminal xylose residues from xylose substituted galacturonic acid residues, for example a [galacturonic acid (3-1) xylose] linkage.

The two galacturonic acid residues between which the polypeptides of the invention cleave may both be xylose substituted, or only one may be xylose substituted or alternatively, neither may be xylose substituted.

Preferably, the polypeptide of the invention is obtainable from a microorganism which possesses a gene encoding an enzyme with endo-xylogalacturonase activity. More preferably the microorganism is a microbial organism, preferably fungal, and optimally a filamentous fungi. Preferred organisms are thus of the genera *Aspergillus*, *Trichoderma*, *Penicillium*, *Acremonium*, *Fusarium*, *Humicola*, *Neurospora*, *Mucor*, *Scytallidium*, *Myceliophtora*, *Thielavia*, *Talaromyces*, *Thermomyces*, *Thermoascus*, *Chaetomium*, *Sporotrichum*, *Corynascus*, *Calcarisporiella* or *Mycelia*. Optionally the organism is of the species from the *Aspergillus niger* Group (as defined by Raper and Fennell, The Genus *Aspergillus*, The Williams & Wilkins Company, Baltimore, pp 293-344, 1965), specifically including but not limited to *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus tubigensis*, *Aspergillus aculeatus*, *Aspergillus foetidus*, *Aspergillus japonicus* or *Aspergillus ficuum*.

In a second aspect, the present invention provides a polynucleotide encoding a polypeptide of the first aspect of the invention. For example the present invention provides a polynucleotide encoding an endo-xylogalacturonase, such an endo-xylogalacturonase whose amino acid sequence is set out in SEQ ID No. 2. The present invention further provides a polynucleotide encoding a polypeptide having

substantial amino acid sequence homology to the amino acid sequence set out in SEQ ID No. 2. Also provided is a polynucleotide selected from:

(a) polynucleotides comprising the nucleotide sequence set out in SEQ ID No. 1, or the complement thereof.

(b) polynucleotides comprising a nucleotide sequence capable of hybridising to the nucleotide sequence set out in SEQ ID No. 1, or a fragment thereof.

(c) polynucleotides comprising a nucleotide sequence capable of hybridising to the complement of the nucleotide sequence set out in SEQ ID No. 1, or a fragment thereof.

(d) polynucleotides comprising a polynucleotide sequence which is degenerate as a result of the genetic code to the polynucleotides defined in (a), (b) or (c).

The term "capable of hybridising" means that the target polynucleotide of the invention can hybridise to the nucleic acid used as a probe (for example the nucleotide sequence set out in SEQ. ID No.1, or a fragment thereof or the complement thereof) at a level significantly above background. The background hybridisation may occur because of, for example, other polynucleotides, such as DNA, present in, for example a cDNA/genomic library being screened. In this event, background implies a level of signal generated by interaction between the probe and a non-specific polynucleotide member of the library which is less than 10 fold, preferably less than 100 fold as intense as the specific interaction observed with the target polynucleotide. The intensity of interaction may be measured, for example, by radiolabelling the probe, e.g. with ^{32}P. Suitable conditions are described below.

Preferably, the polynucleotide of the invention is obtainable from a fungus, in particular a fungus of the genus *Aspergillus*.

The present invention also provides a polynucleotide probe which comprises a fragment of at least 15 nucleotides of a polynucleotide of the invention as described above.

In a third aspect, the invention provides vectors comprising a polynucleotide of the invention, including cloning and expression vectors, and in a fourth aspect methods of growing such vectors in a suitable host cell, for example under conditions in which expression of a protein or polypeptide encoded by a sequence of the invention occurs. Also provided in a fifth aspect are host cells comprising a polynucleotide of the

invention wherein said polynucleotide is heterologous to the genome of said host cell. The term "heterologous to the genome of said host cell" means that the polynucleotide does not naturally occur in the genome of the host cell. Preferably, the host cell is a yeast cell, for example a yeast cell of the genus *Kluyveromyces* or *Saccharomyces* or a fungal cell, for example of the genus *Aspergillus*.

The polypeptides of the invention which possess endo-xylogalacturonase activity may be used in a sixth aspect to treat plant material including plant pulp and plant extracts. For example, they may be used to treat apple pulp and/or raw juice during the production of apple juice. Conveniently the polypeptide of the invention is combined with suitable carriers or diluents including buffers to produce a composition/ enzyme preparation. Thus the present invention provides in a seventh aspect a composition comprising a polypeptide of the invention. The composition may further comprise additional ingredients such as one or more enzymes, for example pectinases, including endo-arabinanase and rhamnogalacturonase, and cellulases and/or xyloglucanases.

The polypeptides and compositions of the invention may therefore be used in a method of processing plant material to degrade or modify the pectin constituents of the cell walls of the plant material. Thus in an eighth aspect, the present invention provides a method of degrading or modifying a plant cell wall which method comprises contacting said plant cell wall with a polypeptide or composition of the invention.

The invention also provides a method of processing a plant material which method comprises contacting said plant material with a polypeptide or composition of the invention to degrade or modify the pectin in said plant material. Preferably said plant material is a plant pulp or plant extract.

In particular, said degradation preferably comprises endo-type cleaving of xylogalacturonan subunits of a pectin component of the plant cell wall. The plant material is preferably a fruit or vegetable pulp or fruit or vegetable extract, for example apple pulp or apple juice.

The present invention further provides a processed plant material obtainable by contacting a plant material with a polypeptide or composition of the invention. Preferably said processed plant material is a fruit or vegetable juice, for example apple juice.

The present invention also provides a method for reducing the viscosity of a plant extract which method comprises contacting the plant extract with a polypeptide or composition of the invention in an amount effective in degrading pectins contained in said plant extract.

Preferred features and characteristics of one aspect of the invention are applicable to another aspect *mutatis mutandis*.

Detailed description of the invention

A. Polynucleotides.

Polynucleotides of the invention may comprise DNA or RNA. They may be single or double stranded. They may also be polynucleotides which include within them synthetic or modified nucleotides.

Polynucleotides of the invention capable of hybridising to the DNA of SEQ ID No. 1 will be generally at least 50%, such as at least 60% or 70% preferably at least 80 or 90% and more preferably at least 95% homologous to the corresponding DNA of SEQ ID No. 1 over a region of at least 20, preferably at least 25 or 30, for instance at least 40, 60 or 100 or more contiguous nucleotides.

It is to be understood that skilled persons may, using routine techniques, make nucleotide substitutions that do not affect the polypeptide sequence encoded by the polynucleotides of the invention to reflect the codon usage of any particular host organism in which the polypeptides of the invention are to be expressed.

Any combination of the above mentioned degrees of homology and minimum sizes may be used to define polynucleotides of the invention, with the more stringent combinations (i.e. higher homology over longer lengths) being preferred. Thus for example a polynucleotide which is at least 80% homologous over 25, preferably 30 nucleotides forms one aspect of the invention, as does a polynucleotide which is at least 90% homologous over 40 nucleotides.

Polynucleotides of the invention may be used to produce a primer, e.g. a PCR primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a revealing label by conventional means using radioactive or non-radioactive labels, or the polynucleotides may be cloned into vectors. Such primers, probes and other fragments will be at least 15, preferably at least 20, for example at least 25, 30 or 40 nucleotides in length, and are also encompassed by the term polynucleotides of the invention as used herein.

Polynucleotides such as a DNA polynucleotide and primers according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques.

In general, primers will be produced by synthetic means, involving a step wise

manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

Longer polynucleotides will generally be produced using recombinant means, for example using a PCR (polymerase chain reaction) cloning techniques. This will involve making a pair of primers (e.g. of about 15-30 nucleotides) to a region of the endo-xylogalacturonase gene which it is desired to clone, bringing the primers into contact with mRNA or cDNA obtained from a fungal, plant or prokaryotic cell, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme recognition sites so that the amplified DNA can be cloned into a suitable cloning vector.

Such techniques may be used to obtain all or part of the endo-xylogalacturonase sequence described herein. Genomic clones containing the endo-xylogalacturonase gene and its introns and promoter regions may also be obtained in an analogous manner, starting with genomic DNA from a fungal, plant or prokaryotic cell.

Although in general the techniques mentioned herein are well known in the art, reference may be made in particular to Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (1989) and Ausubel *et al.*, Current Protocols in Molecular Biology (1995), John Wiley & Sons, Inc.

Polynucleotides which are not 100% homologous to the sequences of the present invention but fall within the scope of the invention can be obtained in a number of ways. Other variants of the endo-xylogalacturonase sequence described herein may be obtained for example by probing genomic DNA libraries made from a range of organisms, for example those discussed as sources of the polypeptides of the invention. In addition, other fungal, plant or prokaryotic homologues of endo-xylogalacturonase may be obtained and such homologues and fragments thereof in general will be capable of hybridising to SEQ ID No. 1. Such sequences may be obtained by probing cDNA libraries or genomic DNA libraries from other species, and probing such libraries with probes comprising all or part of SEQ ID. 1 under conditions of medium to high stringency (for example 0.03M sodium chloride and 0.03M sodium citrate at from about 50°C to about 60°C). Nucleic acid probes comprising all or part of SEQ ID No. 1 may be used to probe cDNA libraries from

other species, such as those described as sources for the polypeptides of the invention .

Species homologues may also be obtained using degenerate PCR which will use primers designed to target sequences within the variants and homologues encoding conserved amino acid sequences. The primers will contain one or more degenerate positions and will be used at stringency conditions lower than those used for cloning sequences with single sequence primers against known sequences.

Alternatively, such polynucleotides may be obtained by site directed mutagenesis of the endo-xylogalacturonase sequences or allelic variants thereof. This may be useful where for example silent codon changes are required to sequences to optimise codon preferences for a particular host cell in which the polynucleotide sequences are being expressed. Other sequence changes may be desired in order to introduce restriction enzyme recognition sites, or to alter the property or function of the polypeptides encoded by the polynucleotides.

The invention further provides double stranded polynucleotides comprising a polynucleotide of the invention and its complement.

Polynucleotides or primers of the invention may carry a revealing label. Suitable labels include radioisotopes such as ^{32}P or ^{35}S, enzyme labels, or other protein labels such as biotin and DIG-hapten. Such labels may be added to polynucleotides or primers of the invention and may be detected using by techniques known *per se*.

The present invention also provides polynucleotides encoding the polypeptides of the invention described below. Since such polynucleotides will be useful as sequences for recombinant production of polypeptides of the invention, it is not necessary for them to be capable of hybridising to the sequence of SEQ ID No. 1, although this will generally be desirable. Otherwise, such polynucleotides may be labelled, used, and made as described above if desired. Polypeptides of the invention are described below.

B. Polypeptides.

Polypeptides of the invention include polypeptides in (preferably substantially) isolated form which comprise the sequence set out in SEQ ID No. 2. Polypeptides further include variants of such sequences, including naturally occurring allelic variants and synthetic variants which are substantially homologous to said polypeptides. In this context, substantial homology is regarded as a sequence which has at least 40%, such

as at least 50%, 60% or 70%, preferably at least 80%, 90% or 95% amino acid homology (identity) over at least 30 amino acids, preferably at least 100 or 200 amino acids, with the sequence of SEQ ID No. 2.

Polypeptides also include those encoding endo-xylogalacturonase homologues, and variants thereof as defined above, from other species. Preferably such homologues will have endo-xylogalacturonase activity as defined above. Preferred species include other fungal species, for example those mentioned as sources of the polypeptides.

Polypeptides of the invention also include fragments of the above mentioned full length polypeptides and variants thereof, including fragments of the sequence set out in SEQ ID No. 2. Suitable fragments will be at least 5, e.g. at least 10, 12, 15 or 20 amino acids in size.

The polypeptides of the invention may also contain one or more (e.g. at least 2, 3, 5, or 10) substitutions, deletions or insertions, including conserved substitutions.

Conserved substitutions may be made according to the following table which indicates conservative substitutions, where amino acids on the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		K R
AROMATIC		H F W Y
OTHER		N Q D E

Polypeptides of the invention may be in a substantially isolated form. It will be understood that the polypeptide may be mixed with carriers or diluents which will not interfere with the intended purpose of the polypeptide and still be regarded as

substantially isolated. A polypeptide of the invention may also be in a substantially purified form, in which case it will generally comprise the polypeptide in a preparation in which more than 90%, e.g. 95%, 98% or 99% of the polypeptide in the preparation is a polypeptide of the invention. Polypeptides of the invention may be modified for example by the addition of histidine residues to assist their purification or by the addition of a signal sequence to promote their secretion from a cell, as discussed below.

Polypeptides of the invention can if necessary be produced by synthetic means although usually they will be made recombinantly as described below.

Particularly preferred polypeptides of the invention include a polypeptide consisting of amino acids 19 to 406 of the amino acid sequence set out in SED. ID No. 2 since this lacks the N-terminal signal peptide which consists of amino acids 1 to 18 of the amino acid sequence of SEQ ID No. 2. The polypeptides and fragments thereof may contain amino acid alterations as defined above.

The use of yeast and fungal host cells is expected to provide for such post-translational modifications (e.g. proteolytic processing, myristilation, glycosylation, truncation, and tyrosine, serine or threonine phosphorylation) as may be needed to confer optimal biological activity on recombinant expression products of the invention.

C. Vectors.

Polynucleotides of the invention can be incorporated into a recombinant replicable vector, for example a cloning or expression vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, the invention provides a method of making polynucleotides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the

vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells are described below in connection with expression vectors.

Expression Vectors and Polypeptide Production

Preferably, a polynucleotide of the invention in a vector is operably linked to a regulatory sequence which is capable of providing for the expression of the coding sequence by the host cell, i.e. the vector is an expression vector. The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A regulatory sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under condition compatible with the control sequences.

The term "regulatory sequences" includes promoters and enhancers and other expression regulation signals.

Enhanced expression of the polynucleotide encoding the polypeptide of the invention may also be achieved by the selection of heterologous regulatory regions, e.g. promoter, secretion leader and terminator regions, which serve to increase expression and, if desired, secretion levels of the protein of interest from the chosen expression host and/or to provide for the inducible control of the expression of the polypeptide of the invention

Aside from the promoter native to the gene encoding the polypeptide of the invention, other promoters may be used to direct expression of the polypeptide of the invention. The promoter may be selected for its efficiency in directing the expression of the polypeptide of the invention in the desired expression host.

In another embodiment, a constitutive promoter may be selected to direct the

expression of the desired polypeptide of the invention. Such an expression construct may provide additional advantages since it circumvents the need to culture the expression hosts on a medium containing an inducing substrate, such as xylogalacturonan.

Examples of strong constitutive and/or inducible promoters which are preferred for use in fungal expression hosts are those which are obtainable from the fungal genes for xylanase (*xln*A), phytase, ATP-synthetase, subunit 9 (*oli*C), triose phosphate isomerase (*tpi*), alcohol dehydrogenase (*Adh*A), α-amylase (*amy*), amyloglucosidase (AG - from the *gla*A gene), acetamidase (*amd*S) and glyceraldehyde-3-phosphate dehydrogenase (*gpd*) promoters.

Examples of strong yeast promoters are those obtainable from the genes for alcohol dehydrogenase, lactase, 3-phosphoglycerate kinase and triosephosphate isomerase.

Examples of strong bacterial promoters are the α-amylase and *SP02* promoters as well as promoters from extracellular protease genes.

Hybrid promoters may also be used to improve inducible regulation of the expression construct.

Often, it is desirable for the polypeptide of the invention to be secreted from the expression host into the culture medium from where the polypeptide of the invention may be more easily recovered. According to the present invention, the polypeptide of the invention's native secretion leader sequence may be used to effect the secretion of the expressed polypeptide of the invention. However, an increase in the expression of the polypeptide of the invention sometimes results in the production of the protein in levels beyond that which the expression host is capable of processing and secreting, creating a bottleneck such that the protein product accumulates within

the cell. Accordingly, the present invention also provides heterologous leader sequences to provide for the most efficient secretion of the polypeptide of the invention from the chosen expression host.

According to the present invention, the secretion leader may be selected on the basis of the desired expression host. A heterologous secretion leader may be chosen which is homologous to the other regulatory regions of the expression construct. For example, the leader of the highly secreted amyloglucosidase (AG) protein may be used in combination with the amyloglucosidase (AG) promoter itself, as well as in combination with other promoters. Hybrid signal sequences may also be used with the context of the present invention.

Examples of preferred heterologous secretion leader sequences are those originating from the fungal amyloglucosidase (AG) gene (*glaA* - both 18 and 24 amino acid versions e.g. from *Aspergillus*), the α-factor gene (yeasts e.g. *Saccharomyces* and *Kluyveromyces*) or the α-amylase gene (*Bacillus*).

Such vectors may be transformed into a suitable host cell as described above to provide for expression of a polypeptide of the invention. Thus, in a further aspect the invention provides a process for preparing polypeptides according to the invention which comprises cultivating a host cell transformed or transfected with an expression vector as described above under conditions to provide for expression by the vector of a coding sequence encoding the polypeptides, and recovering the expressed polypeptides. Suitable host cells include, for example, fungal cells, such as *Aspergillus* and yeast cells, such as yeast cells of the genus *Kluyveromyces* or *Saccharomyces*. Other suitable host cells are discussed below.

The vectors may be for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of the said

polynucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable marker genes. The most suitable selection systems for industrial micro-organisms are those formed by the group of selection markers which do not require a mutation in the host organism. Examples of fungal selection markers are the genes for acetamidase (*amd*S), ATP synthetase, subunit 9 (*oli*C), orotidine-5'-phosphate-decarboxylase (*pyr*A), phleomycin and benomyl resistance (benA). Examples of non-fungal selection markers are the bacterial G418 resistance gene (this may also be used in yeast, but not in fungi), the ampicillin resistance gene (*E. coli*), the neomycin resistance gene (*Bacillus*) and the *E. coli uid*A gene, coding for β-glucuronidase (GUS). Vectors may be used *in vitro*, for example for the production of RNA or used to transfect or transform a host cell.

A further embodiment of the invention provides host cells transformed or transfected with a polynucleotide of the invention. Preferably said polynucleotide is carried in a vector for the replication and expression of said polynucleotides. The cells will be chosen to be compatible with the said vector and may for example be prokaryotic (for example bacterial), fungal, yeast or plant cells.

Bacteria from the genus *Bacillus* are very suitable as heterologous hosts because of their capability to secrete proteins into the culture medium. other bacteria suitable as hosts are those from the genera *Streptomyces* and *Pseudomonas*.

Depending on the nature of the polynucleotide encoding the polypeptide of the invention, and/or the desirability for further processing of the expressed protein, eukaryotic hosts such as yeasts or fungi may be preferred. In general, yeast cells are preferred over fungal cells because they are easier to manipulate. However, some proteins are either poorly secreted from the yeast cell, or in some cases are not processed properly (e.g. hyperglycosylation in yeast). In these instances, a fungal host

organism should be selected.

A heterologous host may also be chosen wherein the polypeptide of the invention is produced in a form which is substantially free from other pectin-degrading enzymes. This may be achieved by choosing a host which does not normally produce such enzymes such as *Kluyveromyces lactis*.

Examples of preferred expression hosts within the scope of the present invention are fungi such as *Aspergillus* species (described in EP-A-184,438 and EP-A-284,603) and *Trichoderma* species; bacteria such as *Bacillus* species (described in EP-A-134,048 and EP-A-253,455), *Streptomyces* species and *Pseudomonas* species; and yeasts such as *Kluyveromyces* species (described in EP-A-096,430 and EP-A-301,670) and *Saccharomyces* species.

Particularly preferred expression hosts may be selected from *Aspergillus niger*, *Aspergillus niger var. tubigenis*, *Aspergillus niger var. awamori*, *Aspergillus aculeatis*, *Aspergillus nidulans*, *Aspergillus orvzae*, *Trichoderma reesei*, *Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus amyloliquefaciens*, *Kluyveromyces lactis* and *Saccharomyces cerevisiae*.

Recombinant production

The invention encompasses processes for the production of the polypeptide of the invention by means of recombinant expression of a DNA sequence encoding the polypeptide. For this purpose the DNA sequence of the invention can be used for gene amplification and/or exchange of expression signals, such as promoters, secretion signal sequences, in order to allow economic production of the polypeptide in a suitable homologous or heterologous host cell. A homologous host cell is herein defined as a host cell which is of the same species or which is a variant within the same species as

the species from which the DNA sequence is derived.

Suitable host cells are preferably microorganisms like bacteria, or more preferably fungi such as yeasts or filamentous fungi. A preferred yeast host cell for the expression of the DNA sequence encoding the polypeptide is of the genera *Saccharomyces*, *Kluyveromyces*, *Hansenula*, *Pichia*, *Yarrowia*, and *Schizosaccharomyces*. More preferably a yeast host cell is selected from the group consisting of the species *Saccharomyces cerevisiae*, *Kluyveromyces lactis* (also known as *Kluyveromyces marxianus* var. *lactis*), *Hansenula polymorpha*, *Pichia pastoris*, *Yarrowia lipolytica*,and *Schizosaccharomyces pombe*.

Most preferred for the expression of the DNA sequence encoding the polypeptide are, however, filamentous fungal host cells. Preferred filamentous fungal host cells are selected from the group consisting of the genera *Aspergillus*, *Trichoderma*, *Fusarium*, *Penicillium*, *Acremonium*, *Neurospora*, *Thermoascus*, *Myceliophtora*, *Sporotrichum*, *Thielavia*, and *Talaromyces*. More preferably a filamentous fungal host cell is of the species *Aspergillus oyzae*, *Aspergillus sojae*, *Aspergillus nidulans*, species from the *Aspergillus niger* Group (as defined by Raper and Fennell, The Genus *Aspergillus*, The Williams & Wilkins Company, Baltimore, pp 293-344, 1965), specifically including but not limited to *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus tubigensis*, *Aspergillus aculeatus*, *Aspergillus foetidus*, *Aspergillus japonicus* and *Aspergillus ficuum*, and further consisting of the species *Trichoderma reesei*, *Fusarium graminearum*, *Penicillium chrysogenum*, *Acremonium alabamense*, *Neurospora crassa*, *Myceliophtora thermophilum*, *Sporotrichum cellulophilum*, and *Thielavia terrestris*.

The DNA sequence encoding the polypeptide is preferably introduced into a suitable host as part of an expression construct in which the DNA sequence is operably

linked to expression signals which are capable of directing expression of the DNA sequence in the host cells. For transformation of the suitable host with the expression construct transformation procedures are available which are well known to the skilled person (see for example: Goosen *et al* for filamentous fungi and Romanos *et al.*, for yeasts). The expression construct can be used for transformation of the host as part of a vector carrying a selectable marker, or the expression construct is co-transformed as a separate molecule together with the vector carrying a selectable marker. Suitable selectable markers which can be used for selection of the transformed host cells are well known to the skilled person (see for example Goosen *et al.*, and Romanos *et al.*). Preferred markers include but are not limited to e.g. versatile marker genes that can be used for transformation of most filamentous fungi and yeasts such as acetamidase genes or cDNAs (the *amdS* genes or cDNAs from *A.nidulans*, *A.oryzae*, or *A.niger*), or genes providing resistance to antibiotics like G418 or hygromycin. Alternatively, more specific selection markers can be used such as auxotrophic markers which require corresponding mutant host strains: e.g. *URA3* (from *S.cerevisiae* or analogous genes from other yeasts), *pyrG* (from *A.nidulans* or *A.niger*) or *argB* (from *A.nidulans* or *A.niger*). In a more preferred embodiment, the selection marker is deleted from the transformed host cell after introduction of the expression construct in accordance with the methods described in EP-A-0 635 574, so as to obtain transformed host cells capable of producing the polypeptide which are free of selection marker genes.

For most filamentous fungi and yeast, the expression construct is preferably integrated in the genome of the host cell in order to obtain stable transformants. However, for certain yeasts also suitable episomal vector systems are available into which the expression construct can be incorporated for stable and high level expression, examples thereof include vectors derived from the 2μ and pKD1 plasmids

of *Saccharomyces* and *Kluyveromyces*, respectively. In case the expression constructs are integrated in the host cells genome, the constructs are either integrated at random loci in the genome, or at predetermined target loci using homolgous recombination, in which case the target loci preferably comprise a highly expressed gene. A highly expressed gene is herein defined as a gene whose mRNA can make up at least 0.5% (w/w) of the total cellular mRNA, e.g. under induced conditions, or alternatively, a gene whose gene product can make up at least 1% (w/w) of the total cellular protein, or, in case of a secreted gene product, can be secreted to a level of at least 0.1 g/l. A number of examples of suitable highly expressed genes is provided herein below.

An expression construct for a given host cell will usually contain the following elements operably linked to each other in a consecutive order from the 5'-end to 3'-end relative to the coding strand of the sequence encoding the polypeptide of the first aspect: (1) a promoter sequence capable of directing transcription of the DNA sequence encoding the polypeptide in the given host cell, (2) optionally, a signal sequence capable of directing secretion of the polypeptide from the given host cell into the culture medium, (3) the DNA sequence encoding a mature and preferably active form of the polypeptide, and preferably also (4) a transcription termination region (terminator) capable of terminating transcription downstream of the DNA sequence encoding the polypeptide.

A variety of promoters capable of directing transcription in the host cells of the invention is available to the skilled person (see for example Goosen *et al.*, and Romanos *et al.*). Preferably the promoter sequence is derived from a highly expressed gene as previously defined. Examples of preferred highly expressed genes from which promoters are preferably derived and/or which are comprised in preferred predetermined target loci for integration of expression constructs, include but are not

limited to genes encoding glycolytic enzymes such as triose-phosphate isomerases (TPI), glyceraldehyde-phosphate dehydrogenases (GAPDH), phosphoglycerate kinases (PGK), pyruvate kinases (PYK), alcohol dehydrogenases (ADH), as well as genes encoding amylases, glucoamylases, xylanases, cellobiohydrolases, ß-galactosidases, alcohol (methanol) oxidases, elongation factors and ribosomal proteins. Specific examples of suitable highly expressed genes include e.g. the *LAC4* gene from *Kluyveromyces* sp., the methanol oxidase genes (*AOX* and *MOX*) from *Hansenula* and *Pichia*, respectively, the glucoamylase (*glaA*) genes from *A.niger* and *A.awamori*, the *A.oryzae* TAKA-amylase gene, the *A.nidulans gpdA* gene and the *T.reesei* cellobiohydrolase genes.

Preferably the polypeptide is produced as a secreted protein in which case the DNA sequence encoding a mature form of the polypeptide in the expression construct is operably linked to a DNA sequence encoding a signal sequence. Preferably the signal sequence is native (homologous) to the DNA sequence encoding the polypeptide. Alternatively the signal sequence is foreign (heterologous) to the DNA sequence encoding the polypeptide, in which case the signal sequence is preferably endogenous to the host cell in which the DNA sequence is expressed. Examples of suitable signal sequences for yeast host cells are the signal sequences derived from yeast α-factor genes. Similarly, a suitable signal sequence for filamentous fungal host cells is e.g. a signal sequence derived from a filamentous fungal (gluco)amylase gene, e.g. the *A.niger glaA* gene.

Downstream of the DNA sequence encoding the polypeptide, the expression construct preferably contains a 3' untranslated region containing one or more transcription termination sites, also referred to as a terminator. The origin of the terminator is less critical. The terminator can e.g. be native to the DNA sequence

encoding the polypeptide. However, preferably a yeast terminator is used in yeast host cells and a filamentous fungal terminator is used in filamentous fungal host cells. More preferably, the terminator is endogenous to the host cell in which the DNA sequence encoding the polypeptide is expressed.

The recombinant host cells according to the invention may be cultured using procedures known in the art. For each combination of a promoter and a host cell, culture condition are available which are conducive to the expression the DNA sequence encoding the polypeptide. After reaching the desired cell density or titre of the polypeptide the culture is stopped and the polypeptide is recovered using known procedures.

Culture of host cells

According to the present invention, the production of the polypeptide of the invention can be effected by the culturing of microbial expression hosts, which have been transformed with one or more polynucleotides of the present invention, in a conventional nutrient fermentation medium.

The fermentation medium can comprise a known culture medium containing a carbon source (e.g. glucose, maltose, molasses, etc.), a nitrogen source (e.g. ammonium sulphate, ammonium nitrate, ammonium chloride, etc.), an organic nitrogen source (e.g. yeast extract, malt extract, peptone, etc.) and inorganic nutrient sources (e.g. phosphate, magnesium, potassium, zinc, iron, etc.). Optionally, an inducer (e.g. apple MHR/ xylogalacturonan) may be included.

The selection of the appropriate medium may be based on the choice of expression hosts and/or based on the regulatory requirements of the expression construct. Such media are well-known to those skilled in the art. The medium may, if

desired, contain additional components favouring the transformed expression hosts over other potentially contaminating microorganisms.

The fermentation can be performed over a period of 0.5-20 days in a batch or fed-batch process suitably at a temperature in the range of between 0 and 45°C and, for example, a pH between 2 and 10. Preferred fermentation conditions are a temperature in the range of between 20 and 37°C and/or a pH between 3 and 9. The appropriate conditions are usually selected based on the choice of the expression host and the protein to be expressed.

After fermentation, the cells can be removed from the fermentation broth by means of centrifugation or filtration. After removal of the cells, the polypeptide of the invention may then be recovered and, if desired, purified and isolated by conventional means.

The polypeptide is typically stably formulated either in liquid or dry form. Typically, the product is made as a composition which will optionally include, for example, stabilising buffers and preservatives. The compositions may also include other enzymes capable of digesting plant material, for example other pectinases such as endo-arabinanase and rhamnogalacturonases. For certain applications, immobilization of the enzyme on a solid matrix may be preferred.

D. Methods of Processing Plant or Pectin-containing Materials

Materials

Plant and pectin-containing materials include plant pulp and plant extracts. In the context of this invention an extract from a plant material is any substance which can be derived from plant material by extraction (mechanical and/or chemical), processing or by other separation techniques. The extract may be juice, nectar, base,

or concentrates made hereof. The plant material may comprise or be from vegetables, e.g., carrots, celery, onions, legumes such as soy beans, or fruit, e.g., pome or seed fruit (apples, pears, quince etc.), grapes, tomatoes, citrus (orange, lemon, lime, mandarin), prunes, cherries, black currants, redcurrants, raspberries, strawberries, cranberries, pineapple and other tropical fruits. According to this invention, apples and apple juice are especially preferred.

Processing Methods

The polypeptides of the invention which possess endo-xylogalacturonase activity may be used to treat plant material including plant pulp and plant extracts. For example, they may be used to treat apple pulp and/or raw juice during the production of apple juice. Conveniently the polypeptide of the invention is combined with suitable carriers or diluents including buffers to produce a composition/ enzyme preparation. The composition may further comprise additional pectinases, for example endo-arabinanase and/or rhamnogalacturonase.

The polypeptides and compositions of the invention may therefore be used in a method of processing plant material to degrade or modify the pectin constituents of the cell walls of the plant material, e.g. see Grassin and Fauquembergue (1996).

Typically, the polypeptides of the invention are used as a composition/ enzyme preparation as described above. The composition will generally be added to plant pulp obtainable by, for example crushing or milling plant material. Incubation of the composition with the plant will typically be carried out for at time of from 10 minutes to 5 hours, such as 30 minutes to 2 hours, preferably for about 1 hour. The processing temperature is preferably 15-25°C, e.g. about 20°C and one can use 10-100g, preferably 30-70g, optimally about 50g of enzyme per ton. The composition may also

include a variety of other plant material-degrading enzymes, for example cellulases and other pectinases. All the enzyme preparations/compositions used may be added sequentially or at the same time to the plant pulp. Depending on the composition of the enzyme preparation the plant material may first be macerated (e.g. to a purée) or liquefied. Using the polypeptides of the invention processing parameters like the yield of the extraction, viscosity of the extract and/or quality of the extract may be improved.

Alternatively, or in addition to the above, a polypeptide of the invention may be added to the raw juice obtained from pressing or liquefying the plant pulp. Treatment of the raw juice will be carried out in a similar manner to the plant pulp. Again, other enzymes may be included. Typical incubation conditions are as described in the previous paragraph. Once the raw juice has been incubated with the polypeptides of the invention, the juice is then filtered to produce the final product.

A composition containing a polypeptide of the invention may also be used during the preparation of fruit or vegetable purees.

The end product of the above processes is typically heat–treated at 85°C for a time of from 1 minute to 1 hour, under conditions to inactivate the polypeptides of the invention.

Due to the highly specific action on pectins the polypeptides of the invention may also be used to prepare pectins with modified characteristics, e.g. modified gelation capacities in specific applications.

The polypeptides of the invention may also be added to animal feeds rich in xylogalacturonan, e.g. soy-containing food, to improve the breakdown of the plant cell wall leading to improved utilisation of the plant nutrients by the animal. The polypeptides of the invention may be added to the feed or silage if pre-soaking or wet

diets are preferred. Advantageously, the polypeptides of the invention may continue to degrade xylogalacturonans in the feed *in vivo*. Fungal derived polypeptides of the invention in particular generally have lower pH optima and are capable of releasing important nutrients in such acidic environments as the stomach of an animal. The invention thus also contemplates (e.g. animal) feeds or foodstuffs comprising one or more polypeptides of the invention.

The polypeptides of the invention may also be used during the production of milk substitutes (or replacers) from soy bean. These milk substitutes can be consumed by both humans and animals. A typical problem during the preparation of these milk substitutes is the high viscosity of the soy bean slurry, resulting in the need for an undesirable dilution of the slurry to a concentration of dry solids of 10 to 15%. An enzyme preparation containing a polypeptide of the invention can be added to, or during the processing of, the slurry, enabling processing at a higher concentration (typically 40 to 50%) dry solids. Also during the preparation of savoury product(s) from soy bean the same problem may occur, and may also be solved by the invention.

The invention will now be described with reference to the following Examples which are intended to be illustrative only and not limiting. In the Figures which accompany the Examples:

Figure 1 is a map of vector pCVlacK.

Figure 2 is a calibration curve for the BCA assay.

Figure 3 is a graph illustrating an HPAEC of xylogalacturonan after degradation by xylogalacturonase.

Figure 4 is a graph illustrating an HPSEC of xylogalacturonan before (thin line) and after (fat line) degradation by xylogalacturonans.

Figure 5 is a graph illustrating a Maldi-ToF mass spectrum of the products of

complete degradation of xylogalacturonan by xylogalacturonase.

Figures 6A-H show degradation of MHR-S by endo-arabinanase, rhamnogalacturonase and xylogalacturonase, separately and in combination.

Figure 7 is a hypothetical structure of the prevailing population of apple MHR having the highest molecular weight (sub unit I is xylogalacturonan, II is stubs of the backbone rich in arabinan side chains, III is rhamnogalacturonase oligomers. Distribution of acetyl groups is not presented but major parts thought to be located within sub unit III).

EXAMPLES

EXAMPLE 1

Construction of the *Aspergillus tubigensis* cDNA expression library

Example 1.1: Construction of an expression vector

Starting vector pGBHSA20 (CBS 997.96) contains the promoter and terminator sequence of the lactase gene (*lac4*) of *K. lactis*, a G418 selection marker and the *E. coli* plasmid pTZ18r for propagation in this host. The *K. lactis* KARSCEN cassette (a gift from Dr. A.A. Winkler, Dept. of Cell Biology and Genetics, Leiden University, The Netherlands) was cloned in a unique *Sma*I site of this vector. The resulting vector was named pCVlacK (Figure 1). The unique *Hin*dIII and *Xho*I sites flanking the *lac4* promoter and terminator, respectively, can be used as cloning sites for cDNA synthesized from *Aspergillus tubigensis* poly(A) RNA.

Example 1.2: Isolation of poly(A) RNA and cDNA synthesis

Aspergillus tubigensis conidia were inoculated in triplo at a density of 10^6 spores/ml in 300 ml of medium containing (per liter): 6 g $NaNO_3$, 0.5 g KCl, 1.5 g KH_2PO_4, 0.5 g $MgSO_4$, pH6.5, 1 ml 1000x Timberlake spore elements (per ml, 50mg EDTA, 22mg $FeSO_4.7H_2O$, 5 mg $MnCl_2.2H_2O$, 22mg $ZnSO_4.7H_2O$, 1.6mg $CuSO_4.5H_2O$, 1.7mg $CoCl_2.6H_2O$, 1.5mg $Na_2MoO_4.2H_2O$, 11mg H_3BO_3, adjusted to pH 6.5) and 10ml 100x Timberlake vitamins (per ml, 0.2mg thiamine-HCl, 0.2mg riboflavin 0.2mg nicotinamide, 1mg pyridoxine-HCl, 0.02mg panthothenic acid, 0.4µm biotin, adjusted to pH 5 to 6), 1 g yeast extract, 5 g Soyoptim (defatted, toasted soy

bean meal from Societé Industrièlle Oleagineux, France). The cultures were incubated in a rotary shaker at 28°C, 150 rpm. The mycelium of one culture was harvested at 10 hours after inoculation, mycelium of the other two cultures was harvested at, respectively, 16 and 24 hours after inoculation. F rom 1 g rinsed and squeezed mycelium total RNA was isolated by the RNAzol method (Cinna/Biotecx). Poly(A) RNA was isolated using Qiagen oligotex columns (Westburg). Equal amounts of poly(A) RNA of time-points 10, 16 and 24 hours, were pooled. cDNA was synthesized using the ZAP-cDNA synthesis kit (Stratagene) with the following modifications: the first-strand synthesis was done with Superscript II reverse transcriptase (GibcoBRL). To 7.5 µg poly(A) RNA, 2 µl linker-primer and RNAse free water was added to a final volume of 28.5 µl. This mixture was incubated for 10 minutes at 70°C and chilled on ice. The following components were added: 10 µl 5x first strand buffer, 5 µl 0.1 M DTT, 3 µl first-strand methyl nucleotide mixture and 1 µl RNAse block. This was incubated for 10 minutes at 25°C, followed by 2 minutes at 42°C. Subsequently, 2.5 µl Superscript II RT (200 U/µl) was added, mixed and incubated for 50 minutes at 42°C. A second modification of the protocol was ligation of a *Hin*dIII adaptor instead of the *Eco*RI adaptor.

The cDNA pool was size separated using a Sephacryl S-500 column. The first fraction eluted from the column did not contain any cDNA, the second and third fraction contained the largest sized cDNA. Subsequent fractions were supposed to contain relatively higher amounts of non-full length cDNA and were of no use for construction of the library. The cDNA of fraction 2 and 3 was ligated in the *Hin*dIII and *Xho*I sites of expression vector pCVlacK (see Figure 1) using the Clontech Ligation Express™ kit. Each ligation mixture was transformed in two batches to electrocompetent *E. coli* XL-Blue MRF'cells. The four transformation suspensions

were plated onto 32 agar plates (LB + 50µg/ml ampicillin). After 16 hours of incubation at 37°C, 7366 transformants were obtained. Bacteria were collected by pouring 2.5 ml LB medium onto a plate and scraping off the cells; 0.5 ml of cell suspension was added to glycerol and stored at -80°C, the remaining 2 ml was used for DNA isolation (Qiagen Spin miniprep kit). In case of a low number of transformants per plate, the 2.5 ml was transferred to a second, third or fourth plate. This yielded 22 pools of about 325 individual transformants. Equal amounts of DNA of each pool were combined for use in *K. lactis* transformation.

Example 1.3: Transformation of the expression library into *K. lactis*

An overnight culture of *K. lactis* strain CBS 2359 grown in YPD (10 g/l yeast extract, 20 g/l Bacto-peptone, 20 g/l glucose) at 30°C was diluted 3000-, 600-, 300- and 100-fold in 150 ml of fresh YPD and incubated for 6 hours at 30°C, 160 rpm in a rotary shaker. The culture with an optical density of 0.7-1.0 was used to prepare electrocompetent cells according to K.N. Faber *et al.* (1994).

Electrocompetent cells were transformed with 1 µg pooled DNA of the *E. coli* library. Electroporation was performed using a Biorad Genepulser with settings at 1.4 kV, 200 Ohm and 25 µF. Transformants were selected on double layer YPD plates (YPD with 20 g/l Bacto-agar): the bottom layer containing 50 µg/ml G418, the top layer was non-selective. 660 µL of transformation mix was plated onto 80 double layer plates. Aliquots of 1.5 and 15 µL were pipetted onto the plates. About 10,000 transformants were obtained.

EXAMPLE 2

Substrate preparation

Example 2.1: Preparation of MHR-S from apples

Modified, saponified hairy regions (MHR-S) from apples were isolated from apples and saponified as described by Schols *et al.* (1990)

Example 2.2: Synthesis of xylogalacturonan from gum tragacanth

5 g of gum tragacanth (Sigma, St. Louis, MO, USA) was suspended in 990 mL of ice-cold distilled water. To this solution 10 mL of an ice-cold 5 M NaOH solution was added. After 24 h, the saponified gum tragacanth (sGT) was dialysed extensively against distilled water at 4°C , and concentrated under reduced pressure to 1 L. For mild acid hydrolysis,

7.65 mL of trifluoric acid (TFA) was added to this sGT solution. Final concentration of TFA in the solution was

0.1 M. The sGT/TFA solution was heated to the boiling point in a microwave, and subsequently incubated in a boiling water bath for 1 h. Finally the hydrolysate was dialyzed extensively against distilled water at 4°C and freeze-dried. This procedure yielded 2.61 g of material which is further referred to as xylogalacturonan.

Example 2.3: Characterization of xylogalacturonan

To determine the sugar composition from both the original gum tragacanth, as well as of the xylogalacturonan, samples were hydrolyzed with 1 M H_2SO_4 (100°C, 3 h) as described by Schols *et al,* 1990, and neutral sugars were converted to their alditol

acetates in order to quantify the individual sugars by Gas Chromatography (GC). The uronic acid content of the hydrolysate was determined colorimetrically using *m*-hydroxybiphenyl (Schols *et al.*, 1990). The sugar composition in mol% of the original gum tragacanth (GT) in comparison with the xylogalacturonan (XG) is shown in the table below.

	Rha	Fuc	Ara	Xyl	Gal	Glc	GalA
GT	2	5	26	17	7	7	36
XG	3	0	1	25	9	6	56

The mild acid hydrolysis has removed effectively arabinosyl and fucosyl residues from this polysaccharide, whereas the GalA:Xyl ratio is more or less unaltered.

The degree of acetyl and methyl esterification of gum tragacanth was estimated by High Pressure Liquid Chromatography (HPLC) as described by Voragen *et al.*, 1986. The degree of methylation and acetylation of gum tragacanth is approximately 75% and 20%, respectively (calculated as mol methyl or actyl groups per mol of GalA). Saponification of the gum removed all methyl and acetyl groups. Molecular weight distribution of the polymers was performed by High Pressure Size Exclusion Chromatochraphy (HPSEC) on three Bio-Gel TSK columns (40XL, 30XL, and 20XL in series as described by Schols *et al.*, 1990. Mild acid hydrolysis of saponified gum tragacanth was accompanied with a decrease in molecular weight, although the resulting material is still high molecular weight. Based on HPSEC elution profiles, xylogalacturonan has an estimated molecular mass of approximately 1,100 kDa, using pullulan reference compounds.

Xylogalacturonan effectively proved to be resistant to enzymatic degradation by all tested endopolygalacturonases and rhamnogalacturonases.

EXAMPLE 3

Screening of the library with the BCA assay

Example 3.1: Growth of the transformants and enzyme preparation

From the about 10 000 *K. lactis* transformants produced as described in Example 1.3, 3,500 individual colonies were picked and transferred to separate wells of multiwell plates. The transformants in the multiwell plates were grown for 48 hours at 30°C in medium I (per 500 mL of H_2O (pH 6.0): Mannitol, 10.00 g; $NH_4H_2PO_4$, 1.50 g; KH_2PO_4, 0.25 g; $(NH_4)_2SO_4$, 0.50 g; $CaCl_2.2H_2O$, 0.01 g; $MgSO_4.7H_2O$, 0.15 g: trace elements H_3BO_3, 375 µg; $CuSO_4.5H_2O$, 40 µg; KI, 75 µg; $MnSO_4.4H_2O$, 300 µg; Na_2MoO_4, 150 µg; $ZnSO_4.7H_2O$, 300 µg; $FeCl_3.6H_2O$, 200 µg) and vitamins Ca-pantothenate, 500 µg; Thiamine, 500 µg; Myo-inositol, 500 µg; Pyridoxine, 500 µg; Nicotinic acid, 500 µg; Biotin, 5 µg) containing 80 ng/mL of the antibiotic G418 and the 35 plates were stored as 15% glycerol stocks.

These transformants were used to inoculate a new set of 35 multi well plates containing 200 µL of the same medium with 80 ng/mL G418 with a replica plater. The *K. lactis* transformants were grown for two days at 30°C in a stove. The cells were precipitated by centrifugation at 3000 rpm in a Hermle zk380 centrifuge.

Example 3.2: Substrate degradation

Carefully, 25 µL of supernatant of each well was pipetted to a new multiwell plate, and 25 µL of a 0.2% solution of substrate (either MHR-S of xylogalacturonan)

in 100 mM NaOAc buffer pH 5.0 was added. After incubation overnight in a stove at 30°C, the increase of reducing carbohydrates was measured with the BCA assay.

Example 3.3: The BCA assay

The BCA assay is based on the reduction of Cu(II) to Cu(I) by reducing carbohydrate mono- and oligomers. A complex is formed of bicinchoninic acid (BCA) and Cu(I). This complex produces an intense purple colour, which can be measured spectrophotometrically. This colour increases with an increasing reducing carbohydrate concentration. The method used in this invention is a modification of the method of Fox and Robyt (1991). However, this is the first time that (a modification of) this method has been used for screening purposes.

The procedure consisted of mixing 10 µL of reducing carbohydrate containing sample from example 3.2 with 90 µL of water and 100 µL of BCA reagent together in a multi well plate. BCA reagent was made freshly each day by mixing two solutions, A and B, 1:1 (v/v) together. Solution A consisted of 54.28 g Na_2CO_3, 24.20 g $NaHCO_3$ and 1.942 g Na_2BCA per liter of destilled water. Solution B consisted of 1.248 g $CuSO_4.5H_2O$ and 1.262 g *L*-serine per liter of destilled water. The plate containing sample, reagent and water was incubated for one hour in a stove at 80°C with a lid on the plate. After cooling the plate for 15 minutes, the absorbance at 550 nm was measured using a multi well plate reader (SLT lab instruments, Austria; model EAR 400). Testlines with galactose show that the assay is linear in the range from 0 to 125 µM galactose (Figure 2).

Transformants that produced in the BCA assay an absorbance 0.1 unit higher than the blank, were checked for xylogalacturonan-degrading capabilities by growing them again and repeating the BCA assay using xylogalacturonan as a substrate. Three

xylogalacturonase producing transformants were found.

EXAMPLE 4

Characterization of xylogalacturonase encoding cDNA

All plasmid inserts of these three transformants were identical, as was found after analysis of the restriction patterns of these inserts. *K. lactis* transformant 27E8, exhibiting xylogalacturonase activity, was used to isolate the pCVlacK expression plasmid by the glass beads method (M.A. Sobanski and Dickinson, 1995). After transformation and propagation of this plasmid in *E. coli*, the cDNA insert was excised from pCVlacK with a *Hin*dIII/*Xho*I digestion. This digestion released a 1.0 and 0.4 kb fragment, due to an internal *Hin*dIII site as appeared from the nucleotide sequence later on. The DNA sequence of the cDNA insert was determined on both strands using 5'- and 3'-specific primers to the *lac4* regulating sequences, and primers based on the cDNA sequence. The DNA sequence of the cDNA insert is presented in SEQ ID No. 1, together with the deduced amino acid sequence. Upstream of the ATG translation startcodon, 20 nucleotides (nt) of 5'-untranslated sequence are present. Downstream the TAA stopcodon 130 nt non-translated sequence followed by the poly-A tail are found. The open reading frame of 1218 nt encodes a protein of 406 amino acids, presented in SEQ ID No. 2. The potential cleavage site of the signal sequence, predicted according to the (-3,-1) rule (G. von Heijne, 1986), is between position 18 and 19.

Comparison of the amino acid sequence to protein databases showed homology to polygalacturonase sequences of prokaryotes, fungi and plants. The percentage similarity was only 30-35% at amino acid level. In addition, 32% similarity to a rhamnogalacturonase of *Aspergillus aculeatus* was found.

EXAMPLE 5

Production of the enzyme

K. lactis transformants expressing the cDNA of endo-xylogalacturonase were transferred from multiwell plate glycerol stocks to reagent tubes: 10 µL glycerol stock was added to 1-2 mL of medium I (see Example 3.1) with 80 ng/mL of G418. These cultures were grown at 30°C in a rotary incubator at 200 rpm for two days and used to inoculate erlenmeyer flasks containing 20 mL of this same medium supplemented with antibiotic. For larger scale production of the enzyme, these cultures were used to inoculate 500 mL of the same medium supplemented with antibiotic in 1 L erlenmeyer flasks. Cells were grown at 30°C in a rotary incubator at 200 rpm for two days. Cultures were centrifuged to precipitate cells, the supernatant was used for the purification.

The crude enzyme preparation (350 mL) was preconcentrated on a Hitrap Q ion-exchange column (Pharmacia Biotech, Sweden) with a flow rate of 0.3 mL/min. Elution was performed on a FPLC system (Pharmacia Biotech, Sweden) with a salt gradient using a 20 mM piperazine pH 5.0 starting buffer (buffer A) and a 0.5 M NaCl in 20 mM piperazine pH 5.0 elution buffer (buffer B). The following gradient was used: to 10% B in 1 minute, to 35% B in 19 minutes, to 100% B in 2 minutes and 100% B for three more minutes. Activity was checked as described in example 3 and active fractions were pooled. They were diluted three times with 20 mM piperazine buffer pH 5.0 and applied on a MiniQ column (Pharmacia Biotech, Sweden). Elution was performed on a Smart system (Pharmacia Biotech, Sweden) with a linear pH gradient from 20 mM piperazine pH 5.0 starting buffer to 10 mM HCl a flow rate of 0.4 mL/min. Active fractions were pooled and investigated using SDS-PAGE. Upon silver staining of the gel one protein band with a molecular mass of approximately 60

kDa was found. The difference with the predicted MW of about 45 kDa, based on the DNA sequence (see Example 4) could be due to protein glycosylation.

EXAMPLE 6

Influences of pH and temperature on enzyme activity

Purified enzyme was used for the characterisation of the enzyme. It was obtained as described in Example 5. Measurements at t=0 were used as blanks. For the determination of the pH stability, the purified enzyme was preincubated without substrate for one hour at a pH range from 2.5 to 8 in McIlvaine buffers. Afterwards the enzyme was incubated with substrate for two hours and the increase in reducing sugars was determined as described in Example 3. The enzyme is stable over a pH range of 3 to 6.

For the determination of the pH and temperature optimum, the purified enzyme was incubated with substrate for two hours at a pH range from 2.5 to 8 or a temperature range from 20 to 80°C. After this, the increase in reducing sugars was determined as described in Example 3. The enzyme has an optimum activity at a temperature of 60°C and at pH.

EXAMPLE 7

Mode of action of the xylogalacturonase

Degradation of xylogalacturonan (modified gum tragacanth, Example 2.2) by the supernatant of the xylogalacturonase-producing *K. lactis* clone was monitored by high performance anion exchange chromatography (HPAEC) and high performance size exclusion chromatography (HPSEC).

HPAEC was performed using a Dionex carbopack PA1 column of a size of 4 x

250 ml. Elution was performed with 0.1 M NaOH (solution A) and 1 M NaOAc in 0.1 NaOH (solution B). The following gradient was used: from 0 to 62% B in 50 minutes, to 100% B in 5 minutes, followed by 100% B during 5 minutes. As can be seen from Figure 3, the enzyme does not produce xylose (expected at a retention time of 5 minutes) or galacturonic acid (expected at a retention time of 15 minutes), not even after 8 hrs incubation. Only oligomers were released, the smallest oligomer being found at a retention time of about 18 minutes: this was a xylose-galacturonic acid dimer. In Figure 3, Bottom line: t=1h, middle line: t=4h, top line: t=8h of incubation.

HPSEC was performed using three columns in series: Bio-Gel TSK 40 (300 x 7.5 mm, from Biorad), Bio-Gel TSK 30 XL (300 x 7.5 mm, from Biorad) and TSKGel G 2500 P XL (300 x 7.8 mm, from TosoHaas). Figure 4 shows that the high molecular weight fraction of xylogalacturonan (left in the picture) is rapidly degraded. In Figure 4, the thin grey line represents the polymer before degradation and the thicker black line represents the polymer after degradation.

When the degradation products were monitored by Maldi-ToF mass spectrometry the following products were identified (Figure 5):

Oligomers	**Composition**	**Peak numbers**	**Molecular weight (Da)**
dimer	galAxyl	1	349.1
trimer	galA2xyl	2	525.1
tetramers	galA2xyl2, galA3xyl	3a, 3b	657.3, 701.3
pentamers	galA3xyl2, galA4xyl1	4a, 4b	833.4, 877.4
hexamer	galA4xyl2	5	1009.5
heptamers	galA4xyl3, galA3xyl4	6a, 6b	1141.6, 1185.5
octamer	galA6xyl2	7	1361.5

These results show that xylogalacturonan is degraded by the xylogalacturonase in an endo-fashion.

Upon incubation of the supernatant of the xylogalacturonase producing *K.lactis* transformant with polygalacturonic acid no degradation of the substrate was observed.

EXAMPLE 8

Complete degradation of MHR-S by xylogalacturonase in combination with other enzymes

To study the degradation of MHR-s 200 µL of a 0.3% MHR-S solution in 50 mM NaOAc buffer pH 5.0, was incubated with either 5 µL of the purified xylogalacturonase, or 5 µL of endo-arabinanase (Beldman *et al.*, 1993), or 5 µL of rhamnogalacturonase (H.A. Schols *et al.*, 1990a), or with combinations of these enzyme, added sequentially or at the same time. MHR-S without enzyme was used as a blank.

Degradation of MHR-S was monitored with HPSEC, as described in Example 7. The results are shown in Figure 6 where thin lines represent the blanks and thick lines represent the incubations with enzyme. Incubations were with A: endo-arabinanase, B: xylogalacturonase, C: rhamnogalacturonase, D: endo-arabinanase and xylogalacturonase sequentially, E: endo-arabinanase and xylogalacturonase combined, F: endo-arabinanase and rhamnogalacturonase sequentially, G: endo-arabinanase rhamnogalacturonase and xylogalacturonase sequentially, H: endo-arabinanase rhamnogalacturonase and xylogalacturonase combined.

Figure 6B shows that xylogalacturonase is able to degrade MHR-S: a small shift to lower molecular weight material can be observed. Also the enzymes endo-arabinanase (Figure 6A) and rhamnogalacturonase (Figure 6C) cause some shift in

molecular weight. However, somewhat better results are obtained by combining two different enzymes in one incubation (Figures 6D – endo-arabinanase and endo-xylogalacturonase sequentially; 6E - endo-arabinanase and endo-xylogalacturonase combined and 6F -endo-arabinanase and endo-rhamnogalacturonase sequentially). Almost complete degradation of the high molecular weight material was possible when the three enzymes were added either sequentially (Figure 6G) or combined (Figure 6H).

REFERENCES

H.A. Schols, M.A. Posthumus, A.G.J. Voragen (1990) *Carbohydrate Research* 206: 117-129

H.A. Schols, C.C., J.M. Geraerds, M.J.F. Searle van Leeuwen, F.J.M. Kormelink, A.G.J. Voragen (1990a) *Carbohydrate Research* 206:105-115

A.G.J. Voragen, H.A. Schols and W. Pilnik, (1986) *Food Hydrocoll.* 1: 65-70

G. Beldman, M.J.F. Searle van Leeuwen, G.A. de Ruiter, H.A. Siliha, A.G.J. Voragen (1993) *Carbohydrate Polymers* 20: 159-168

G. Beldman, L.A.M. van den Broek, H.A. Schols, M.J.F Searle-van Leeuwen, K.M.J van Laere and A.G.J. Voragen (1996) *Biotechnology Letters* 18: 707-712

J.D. Fox, J.F. Robyt (1991) *Analytical Biochemistry* 195: 93-96

M.A. Sobanski and Dickinson (1995) *Biotechnology Techniques* 9: 225-230)

G. von Heijne (1986) *Nucleic Acids Research* 14: 4683-4690)

K.N. Faber, P. Haima, W. Harder, M. Veenhuis, G. AB (1994), *Current Genetics* 25: 305-310.

K. Blondeau, O. Boutur, G. Moulin, P. Galzy (1993) *Biotechnology Techniques* 7(8): 609-614.

C. Grassin and P. Fauquembergue (1996) *Pectins and Pectinases, Progress in Biotechnology* 14: 453-462

Goosen *et al.*, 1992, "Transformation and Gene Manipulation in filamentous fungi: an overview" In: Handbook of Applied Mycology" Vol. 4: "Fungal Biotechnology", D.K.

Arora, R.P. Elander and K.G. Mukerji, eds., Marcel Deker In., New York, pp 151-195
Romanos *et al*, Yeast 8: 423-488 (1992)

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Gist-brocades B.V.
(B) STREET: Wateringseweg 1
(C) CITY: Delft
(E) COUNTRY: the Netherlands
(F) POSTAL CODE (ZIP): 2611 XT

(ii) TITLE OF INVENTION: 2895

(iii) NUMBER OF SEQUENCES: 2

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(v) CURRENT APPLICATION DATA:
APPLICATION NUMBER:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1602 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Aspergillus tubigensis

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 98..1318

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
GCTTGTGTTT CTTAGGAGAA TTATTATTCT TTTGTTATGT TGCGCTTGTA GTTGGAAAAG
60

GTGAAGAGAC AAAGCTTGAA TTCCGAAATC GCTCATC ATG GCG CTA TAT CGT AAC
115
                                          Met Ala Leu Tyr Arg Asn
                                            1               5

CTC TAC CTT CTG GCC AGC CTT GGG CTA AGC AGT GCT GCT CCC TCC AAG
163
Leu Tyr Leu Leu Ala Ser Leu Gly Leu Ser Ser Ala Ala Pro Ser Lys
             10                  15                  20

GTC CAG CGA GCC CCG GAT TCT TCC ATT CAT GCT CGC GCT GTC TGT ACC
211
Val Gln Arg Ala Pro Asp Ser Ser Ile His Ala Arg Ala Val Cys Thr
         25                  30                  35
```

```
CCG ACC GCA GGA GGC GAT TCG TCC ACC GAC GAT GTC CCC GCC ATC ACC
259
Pro Thr Ala Gly Gly Asp Ser Ser Thr Asp Asp Val Pro Ala Ile Thr
 40                  45                  50

GAG GCC CTC AGC TCG TGC GGA AAT GGT GGC ACC ATC GTC TTC CCC GAG
307
Glu Ala Leu Ser Ser Cys Gly Asn Gly Gly Thr Ile Val Phe Pro Glu
 55                  60                  65                  70

GGC AGC ACC TAC TAC CTC AAC AGT GTG CTG GAC TTG GGC AGC TGC AGT
355
Gly Ser Thr Tyr Tyr Leu Asn Ser Val Leu Asp Leu Gly Ser Cys Ser
                 75                  80                  85

GAT TGC GAC ATC CAG GTG GAA GGT CTT CTG AAG TTC GCC AGC GAT ACC
403
Asp Cys Asp Ile Gln Val Glu Gly Leu Leu Lys Phe Ala Ser Asp Thr
             90                  95                 100

GAT TAC TGG AGC GGT CGC ACT GCC ATG ATC AGT GTT TCC AAT GTA GAT
451
Asp Tyr Trp Ser Gly Arg Thr Ala Met Ile Ser Val Ser Asn Val Asp
        105                 110                 115

GGT TTG AAG CTG CGC TCA TTG ACT GGA TCT GGT GTC ATT GAT GGC AAT
499
Gly Leu Lys Leu Arg Ser Leu Thr Gly Ser Gly Val Ile Asp Gly Asn
    120                 125                 130

GGC CAG GAT GCG TGG GAT CTC TTT GCT TCG GAC AGT AGT TAC TCA CGC
547
Gly Gln Asp Ala Trp Asp Leu Phe Ala Ser Asp Ser Ser Tyr Ser Arg
135                 140                 145                 150

CCG ACG CTC TTG TAC ATC ACT GGC GGC AGC AAC CTA GAA ATC TCC GGG
595
Pro Thr Leu Leu Tyr Ile Thr Gly Gly Ser Asn Leu Glu Ile Ser Gly
                155                 160                 165

CTG CGT CAA AAG AAT CCA CCT AAC GTG TTC AAC TCG GTC AAG GGT GGC
643
Leu Arg Gln Lys Asn Pro Pro Asn Val Phe Asn Ser Val Lys Gly Gly
            170                 175                 180

GCC ACT AAT GTC GTC TTC TCC AAC CTG AAG ATG GAT GCC AAC TCC AAG
691
Ala Thr Asn Val Val Phe Ser Asn Leu Lys Met Asp Ala Asn Ser Lys
        185                 190                 195

TCG GAC AAT CCG CCC AAG AAC ACT GAT GGG TTC GAC ATT GGC GAG AGT
739
Ser Asp Asn Pro Pro Lys Asn Thr Asp Gly Phe Asp Ile Gly Glu Ser
    200                 205                 210

ACC TAT GTG ACC ATC ACC GAG GTC ACC GTA GTC AAC GAT GAC GAC TGT
787
Thr Tyr Val Thr Ile Thr Glu Val Thr Val Val Asn Asp Asp Asp Cys
215                 220                 225                 230

GTC GCC TTC AAG CCC AGT TCC AAC TAC GTG ACA GTG GAC ACG ATC AGC
835
Val Ala Phe Lys Pro Ser Ser Asn Tyr Val Thr Val Asp Thr Ile Ser
                235                 240                 245

TGC ACC GGC TCC CAT GGA ATT TCC GTG GGA TCA TTA GGA AAG TCG AGC
883
Cys Thr Gly Ser His Gly Ile Ser Val Gly Ser Leu Gly Lys Ser Ser
```

```
                250                 255                 260

GAC GAC TCG GTC AAG AAC ATT TAT GTC ACG GGC GCA ACT ATG ATC AAC
931
Asp Asp Ser Val Lys Asn Ile Tyr Val Thr Gly Ala Thr Met Ile Asn
     265                 270                 275

TCC ACC AAA GCC GCC GGG ATC AAG ACT TAT CCG AGT GGA GGC GAC CAC
979
Ser Thr Lys Ala Ala Gly Ile Lys Thr Tyr Pro Ser Gly Gly Asp His
280                 285                 290

GGT ACC TCC ACG GTC AGC AAT GTG ACC TTC AAC GAT TTC ACT GTG GAC
1027
Gly Thr Ser Thr Val Ser Asn Val Thr Phe Asn Asp Phe Thr Val Asp
295                 300                 305                 310

AAC TCC GAC TAT GCC TTC CAG ATC CAG AGC TGC TAT GGC GAG GAC GAT
1075
Asn Ser Asp Tyr Ala Phe Gln Ile Gln Ser Cys Tyr Gly Glu Asp Asp
                315                 320                 325

GAC TAT TGC GAG GAA AAC CCG GGC AAC GCC AAA CTG ACT GAT ATA GTC
1123
Asp Tyr Cys Glu Glu Asn Pro Gly Asn Ala Lys Leu Thr Asp Ile Val
          330                 335                 340

GTG TCA AGC TTC AGT GGG ACA ACC AGT GAC AAG TAC GAT CCG GTC GTG
1171
Val Ser Ser Phe Ser Gly Thr Thr Ser Asp Lys Tyr Asp Pro Val Val
     345                 350                 355

GCC AAC CTC GAC TGC GGT GCG GAT GGA ACT TGT GGC ATC TCC ATC AGT
1219
Ala Asn Leu Asp Cys Gly Ala Asp Gly Thr Cys Gly Ile Ser Ile Ser
360                 365                 370

GGG TTC GAT GTC AAG GCG CCA TCG GGC AAG TCT GAA GTG TTG TGC GCC
1267
Gly Phe Asp Val Lys Ala Pro Ser Gly Lys Ser Glu Val Leu Cys Ala
375                 380                 385                 390

AAC ACC CCG TCT GAT TTG GGC GTC ACT TGC ACT TCG GGG GCT TCG GGC
1315
Asn Thr Pro Ser Asp Leu Gly Val Thr Cys Thr Ser Gly Ala Ser Gly
                395                 400                 405

TAAATAGCTT TGGCCGGGTT GCTTTCTGAA TCCACTGAGT GGAGGTCTTC TTCGGGTTTG
1375

ATATTTTGTA TGGTCGTGTG TATAGCAGAA TGTGACAATA GAATTAGTGA AATTGCCATT
1435

CTTTTCGAAA GACAAAAAAA AAAAAAAAAA AAAAAAAAAA ACTCGAGAAT TTATACTTAG
1495

ATAAGTATGT ACTTACAGGT ATATTTCTAT GAGATACTGA TGTATACATG CATGATAATA
1555

TTTAAACGGT TATTAGTGCC GATTGTCTTG TGCGATAATG ACGTTCC
1602
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 406 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ala Leu Tyr Arg Asn Leu Tyr Leu Leu Ala Ser Leu Gly Leu Ser
 1               5                  10                  15

Ser Ala Ala Pro Ser Lys Val Gln Arg Ala Pro Asp Ser Ser Ile His
            20                  25                  30

Ala Arg Ala Val Cys Thr Pro Thr Ala Gly Gly Asp Ser Ser Thr Asp
        35                  40                  45

Asp Val Pro Ala Ile Thr Glu Ala Leu Ser Ser Cys Gly Asn Gly Gly
    50                  55                  60

Thr Ile Val Phe Pro Glu Gly Ser Thr Tyr Tyr Leu Asn Ser Val Leu
65                  70                  75                  80

Asp Leu Gly Ser Cys Ser Asp Cys Asp Ile Gln Val Glu Gly Leu Leu
                85                  90                  95

Lys Phe Ala Ser Asp Thr Asp Tyr Trp Ser Gly Arg Thr Ala Met Ile
            100                 105                 110

Ser Val Ser Asn Val Asp Gly Leu Lys Leu Arg Ser Leu Thr Gly Ser
        115                 120                 125

Gly Val Ile Asp Gly Asn Gly Gln Asp Ala Trp Asp Leu Phe Ala Ser
    130                 135                 140

Asp Ser Ser Tyr Ser Arg Pro Thr Leu Leu Tyr Ile Thr Gly Gly Ser
145                 150                 155                 160

Asn Leu Glu Ile Ser Gly Leu Arg Gln Lys Asn Pro Pro Asn Val Phe
                165                 170                 175

Asn Ser Val Lys Gly Gly Ala Thr Asn Val Val Phe Ser Asn Leu Lys
            180                 185                 190

Met Asp Ala Asn Ser Lys Ser Asp Asn Pro Pro Lys Asn Thr Asp Gly
        195                 200                 205

Phe Asp Ile Gly Glu Ser Thr Tyr Val Thr Ile Thr Glu Val Thr Val
    210                 215                 220

Val Asn Asp Asp Asp Cys Val Ala Phe Lys Pro Ser Ser Asn Tyr Val
225                 230                 235                 240

Thr Val Asp Thr Ile Ser Cys Thr Gly Ser His Gly Ile Ser Val Gly
                245                 250                 255

Ser Leu Gly Lys Ser Ser Asp Asp Ser Val Lys Asn Ile Tyr Val Thr
            260                 265                 270

Gly Ala Thr Met Ile Asn Ser Thr Lys Ala Ala Gly Ile Lys Thr Tyr
        275                 280                 285

Pro Ser Gly Gly Asp His Gly Thr Ser Thr Val Ser Asn Val Thr Phe
    290                 295                 300

Asn Asp Phe Thr Val Asp Asn Ser Asp Tyr Ala Phe Gln Ile Gln Ser
305                 310                 315                 320

Cys Tyr Gly Glu Asp Asp Asp Tyr Cys Glu Glu Asn Pro Gly Asn Ala
                325                 330                 335
```

Lys Leu Thr Asp Ile Val Val Ser Ser Phe Ser Gly Thr Thr Ser Asp
340 345 350

Lys Tyr Asp Pro Val Val Ala Asn Leu Asp Cys Gly Ala Asp Gly Thr
355 360 365

Cys Gly Ile Ser Ile Ser Gly Phe Asp Val Lys Ala Pro Ser Gly Lys
370 375 380

Ser Glu Val Leu Cys Ala Asn Thr Pro Ser Asp Leu Gly Val Thr Cys
385 390 395 400

Thr Ser Gly Ala Ser Gly
405

(3)

CLAIMS

1. A polypeptide which possesses endo-xylogalacturonase activity.

2. An isolated polypeptide having endo-xylogalacturonase activity characterised in that said polypeptide is obtainable from a fungus which encodes a protein that possesses endo-xylogalacturonase activity.

3. A polypeptide according to claim 2 wherein said fungus is of the genus *Aspergillus*.

4. A polypeptide which comprises the sequence set out in SEQ ID No. 2, or a polypeptide substantially homologous thereto, or a fragment of the polypeptide of SEQ ID No. 2.

5. A polypeptide which comprises a fragment of a polypeptide as defined in claim 1 or claim 2, said fragment having at least 5 amino acids.

6. A polypeptide according to claim 5 which comprises amino acids 19 to 406 of the amino acid sequence set out in SEQ ID No. 2.

7. A polynucleotide encoding a polypeptide according to any one of the preceding claims.

8. A polynucleotide selected from:

(a) polynucleotides comprising the nucleotide sequence set out in SEQ ID No. 1, or the complement thereof.

(b) polynucleotides comprising a nucleotide sequence capable of hybridising to the nucleotide sequence set out in SEQ ID No. 1, or a fragment thereof.

(c) polynucleotides comprising a nucleotide sequence capable of hybridising to the complement of the nucleotide sequence set out in SEQ ID No. 1, or a fragment thereof.

(d) polynucleotides comprising a polynucleotide sequence which is degenerate as a result of the genetic code to the polynucleotides defined in (a), (b) or (c).

9. An isolated polynucleotide according to claim 7 or claim 8 obtainable from a fungus.

10. A polynucleotide according to claim 9 wherein said fungus is of the genus *Aspergillus*.

11. A polynucleotide probe which comprises a fragment of at least 15 nucleotides of a polynucleotide as defined in claim 7 or 8.

12. A vector comprising a polynucleotide as defined in any one of claims 7 to 11.

13. An expression vector comprising a polynucleotide as defined in any one of claims 7 to 10, operably linked to regulatory sequences capable of directing expression of said polynucleotide in a host cell.

14. An expression vector according to claim 13 wherein said host cell is a yeast cell.

15. A host cell transformed or transfected with a vector according to any one of claims 12 to 14.

16. A host cell comprising a polynucleotide according to any one of claims 7 to 10 wherein said polynucleotide is heterologous to the genome of said host cell.

17. A host cell according to claim 15 or claim 16 which is a yeast cell.

18. A method for producing a polypeptide according to any one of claims 1 to 6 which comprises incubating a host cell according to any one of claims 15 to 17 under conditions which allow the expression of said polypeptide, and optionally purifying said polypeptide.

19. A host cell comprising a polypeptide according to any one of claims 1 to 6 wherein said polypeptide is heterologous to said host cell.

20. A composition comprising a polypeptide according to any one of claims 1 to 6.

21. A composition according to claim 20 which further comprises a polypeptide having endo-arabinanase activity.

22. A method of degrading or modifying a plant cell wall which method comprises contacting said plant cell wall with a polypeptide according to any one of claims 1 to 6 or a composition according to claim 20 or claim 21.

23. A method according to claim 22 wherein said degradation comprises endo-type cleaving of xylogalacturonan subunits of a pectin component of said cell wall.

24. A method of processing a plant material which method comprises contacting said plant material with a polypeptide according to any one of claims 1 to 6 or composition according to claim 10 or claim 11 to degrade or modify the pectin in

said plant material.

25. A method according to claim 24 wherein said plant pulp or plant extract is fruit or vegetable pulp or fruit or vegetable extract.

26. A method according to claim 25 wherein said fruit extract is apple juice.

27. A processed plant material obtainable by contacting a plant material with a polypeptide according to any one of claims 1 to 6 or a composition according to claim 20 or claim 21, or which results from a method according to claim 25 or claim 26.

28. A processed plant material according to claim 27 which is a fruit or vegetable juice.

29. A method for reducing the viscosity of a plant extract which method comprises contacting the plant extract with a polypeptide according to any one of claims 1 to 6 or a composition according to claim 20 or claim 21 in an amount and under conditions effective to degrade pectin contained in said plant extract.

30. Use of a polypeptide according to any one of claims 1 to 6 or a composition according to claim 20 or claim 21 in a method of degrading or modifying a plant cell wall.

31. Use according to claim 30 wherein said degradation comprises endo-type cleaving xylogalacturonan substituents of the pectin components of said cell wall.

32. Use of a polypeptide according to any one of claims 1 to 6 or a composition according to claim 20 or claim 21 in a method of processing plant pulp or plant extract which method comprises incubating said plant pulp or plant extract with said polypeptide or composition to degrade the pectin in said plant pulp or plant extract.

33. An (animal) feed or foodstuff comprising a polypeptide according to any one of the claims 1 to 6.

(4)

ABSTRACT

ENDO-XYLOGALACTURONASE AND HOMOLOGUES THEREOF

A novel polypeptide which has endo-xylogalacturonase activity and polynucleotides encoding the endo-xylogalacturonase are disclosed, as are vectors and hosts. The endo-xylogalacturonase can be used in methods of processing plant materials to produce fruit juices and other plant extracts.

(5)

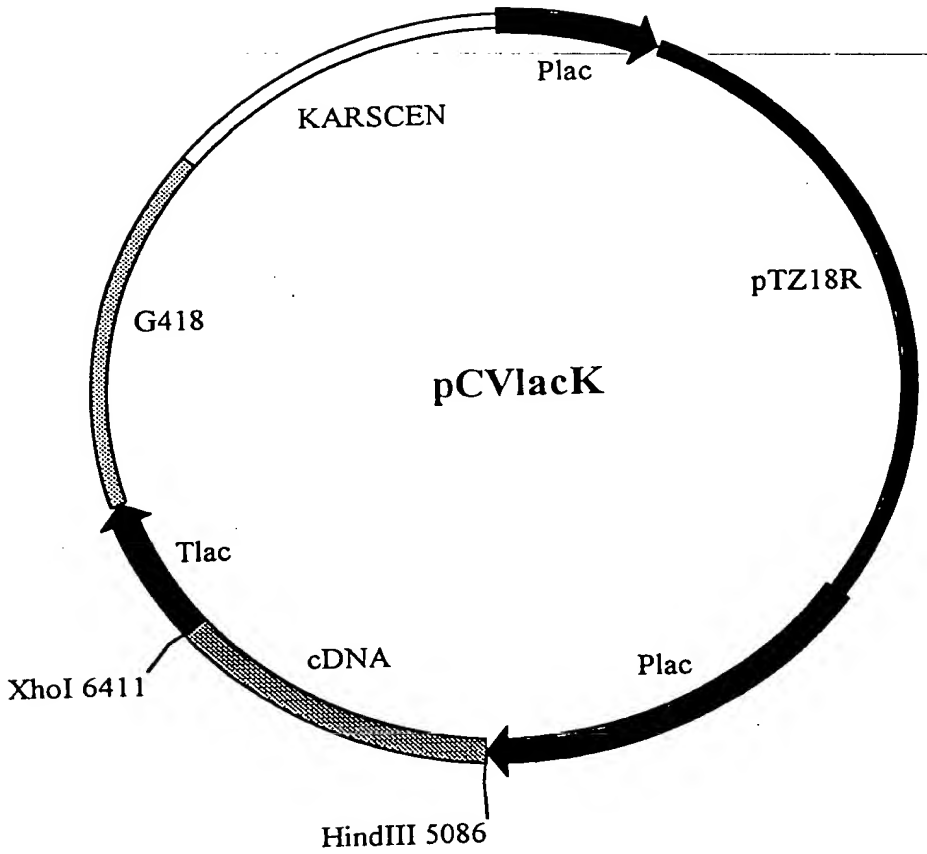

Plac
KARSCEN
pTZ18R
G418
pCVlacK
Tlac
cDNA
Plac
XhoI 6411
HindIII 5086

FIGURE 1

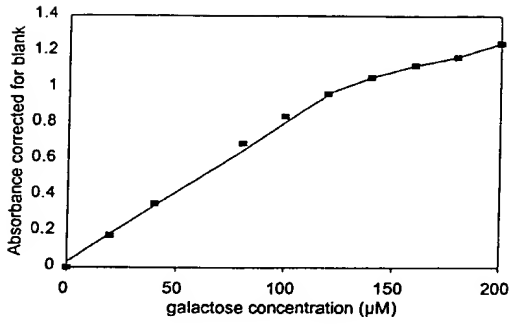


FIGURE 2

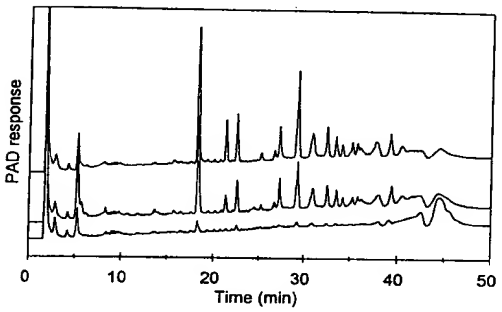


FIGURE 3

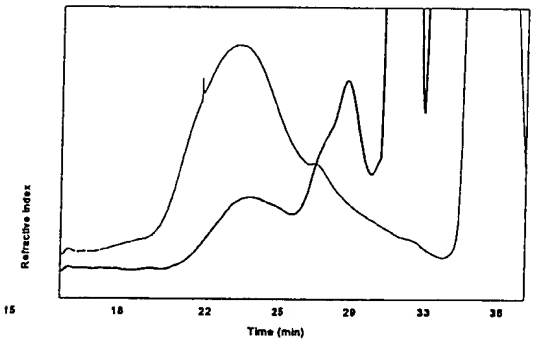


FIGURE 4

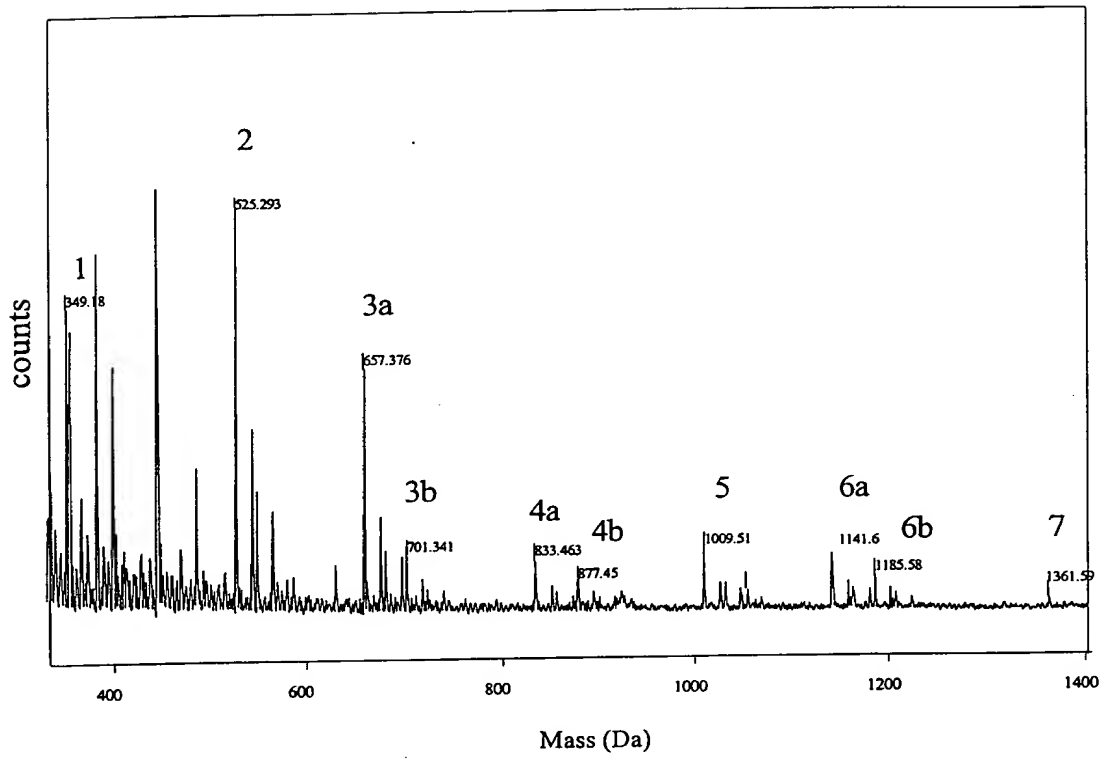
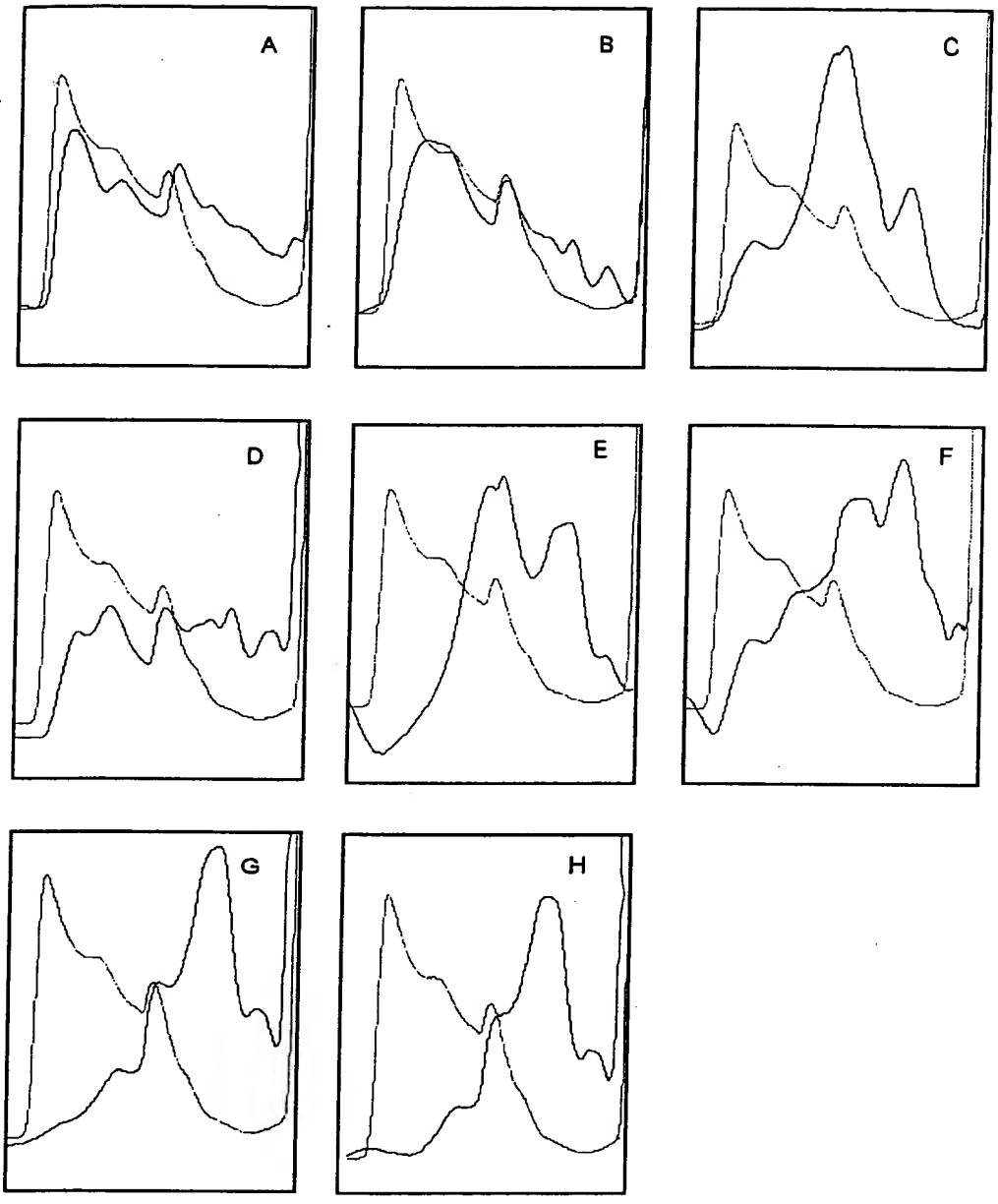

counts
1
349.18
2
525.293
3a
657.376
3b
701.341
4a
833.463
4b
877.45
5
1009.51
6a
1141.6
6b
1185.58
7
1361.59
400
600
800
1000
1200
1400
Mass (Da)

FIGURE 5



FIGURES 6A - 6H

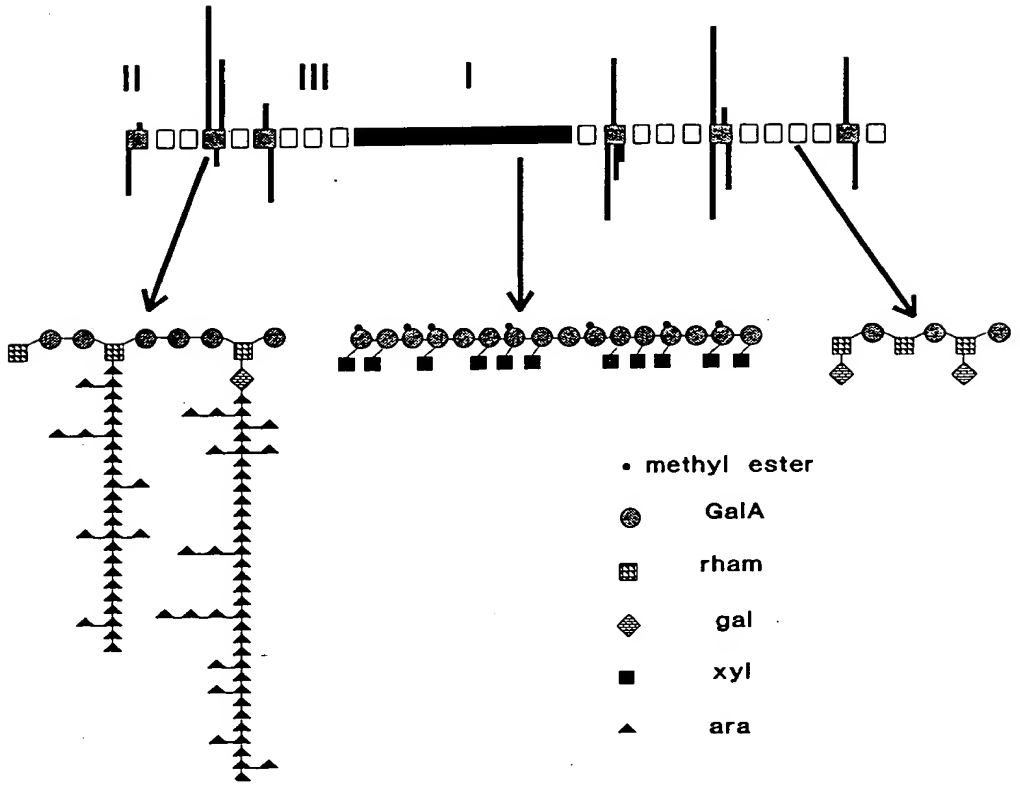


FIGURE 7